Run-Time Support for Integrated Power and Resilience Management

Dorian Arnold[*], Patrick Bridges[†], Kurt Ferreria[†], David K. Lowenthal[‡], Martin Schulz[§]

Introduction

Two of the most critical exascale computing challenges are power and resilience. Power is a significant concern because the DOE exascale power budget for exascale systems is 20 megawatts, yet current 10-20 Petaflop systems are already nearing 10 megawatts. Resilience is already a major challenge for petascale applications because of increasing system failure rates and I/O rates, which will only become worse in exascale systems. If both of these challenges are not sufficiently addressed, it becomes impossible to realize exascale computing systems for any viable system budget.

Approach

At exascale, the scientific computing community must view power and resilience management as performance and capability problems that the runtime and system software must address. In other words, researchers must develop run-time systems that control power and resilience so that the application performance goals are met. To achieve this, research must be carried out to **automatically infer** and **dynamically maintain** techniques to constrain power and choose resilience mechanisms for a variety of exascale applications.

The steps that should be taken are as follows:

- Evaluate applications for their sensitivity to power. This includes (1) evaluating the various demands on different subsystems (e.g., processor and memory) and (2) evaluating how scalability affects the choice of the target number of nodes and the power allocated to them.
- Evaluate applications for their sensitivity to resilience. This includes (1) evaluating different resilience techniques (e.g., checkpointing versus replication), (2) evaluating data and communication characteristics and (3) evaluating failure distributions of the underlying hardware.

This work was partially performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000

[*]Dept. of Computer Science, University of New Mexico

[†]Sandia National Laboratories

[‡]Dept. of Computer Science, The University of Arizona

[§]Lawrence Livermore National Laboratory

- Create performance models to be used by the run-time system to manage system power and resilience settings. The large search space of power and resilience settings requires a model to allow the run-time system to operate effectively.
- Extend the performance model to handle important interactions between power and resilience. For example, given a limit on power, the run-time system could choose to reduce power to nodes, use fewer nodes/cores, or disable replicas/reduce checkpoint frequency. Another example is that replicas could be chosen according to temperature of different nodes or racks, as hotter nodes are more likely to fail. Communication between the power and resilience subsystems will clearly be necessary.
- Create the run-time system that will manage and control power and resilience. One of its duties will be collecting performance information to be used in the performance model. Another will be to manage adaptation, as conditions are likely to change during program execution. For example, failure rates could change due to different job allocations on nearby nodes.

Related Work

There has been a significant amount of work in power and energy reduction for high-performance computing [1, 6]. Also, many have studied checkpointing, including uncoordinated checkpointing [5] and scalable checkpoint/restart (SCR) [8]. Also, several have created analytic models to predict or to understand energy consumption in the context of scalability [9, 4, 7]. There has also been work on modeling the run time of an application for different failure rates and checkpoint/recovery rates [2], as well as work on mathematical models of the impact of replication on application mean time to interrupt and hardware efficiency [3].

Assessment

In this section we assess our proposed research.

- *Challenges addressed.* Power and resilience are two of the critical problems for achieving exascale.
- *Maturity.* There has been significant work in using DVFS to save energy in HPC programs, as well as in using checkpointing for resilience. These two general areas have been reasonably successful. There is also recent work on replication as a replacement for checkpointing. These techniques can be leveraged in the approach suggested above.
- *Uniqueness.* Many are interested in power and resilience in other domains. Exascale computing is unique in that both are simultaneous concerns, and the metric to optimize is performance.
- *Novelty.* We do not know of any work in integrating power and resilience in high-performance computing.
- *Applicability.* This line of research could be useful in all high-performance computing areas, from petascale on up.
- *Effort.* We estimate that this effort would require multiple years of effort. The effort ideally would coordinate research groups in power, resilience, and run-time system construction.

References

[1] K. W. Cameron, X. Feng, and R. Ge. Performance-constrained distributed DVS scheduling for scientific applications on power-aware clusters. In *Supercomputing*, Seattle, Washington, Nov. 2005.

[2] J. T. Daly. A higher order estimate of the optimum checkpoint interval for restart dumps. *Future Gener. Comput. Syst.*, 22(3):303–312, 2006.

[3] K. B. Ferreira, R. Riesen, P. G. Bridges, D. Arnold, J. Stearley, J. H. Laros, R. A. Oldfield, K. Pedretti, and R. Brightwell. Evaluating the viability of process replication reliability for exascale systems. In *Supercomputing*, November 2011.

[4] R. Ge, X. Feng, W. Feng, and K. W. Cameron. CPU Miser: A performance-directed, run-time system for power-aware clusters. In *Proceedings of the 2007 International Conference on Parallel Processing*, Xi'An, China, 2007.

[5] A. Guermouche, T. Ropars, E. Brunet, M. Snir, and F. Cappello. Uncoordinated checkpointing without domino effect for send-deterministic MPI applications. In *IPDPS*, May 2011.

[6] C.-H. Hsu and W.-C. Feng. A power-aware run-time system for high-performance computing. In *Supercomputing*, Nov. 2005.

[7] J. Li and J. F. Martìnez. Dynamic power-performance adaptation of parallel computation on chip multiprocessors. In *12th International Symposium on High-Performance Computer Architecture*, Austin, Texas, Feb. 2006.

[8] A. Moody, G. Bronevetsky, K. Mohror, and B. R. d. Supinski. Design, modeling, and evaluation of a scalable multi-level checkpointing system. In *Supercomputing*, Nov. 2010.

[9] R. Springer, D. K. Lowenthal, B. Rountree, and V. W. Freeh. Minimizing execution time in MPI programs on an energy-constrained, power-scalable cluster. In *ACM Symposium on Principles and Practice of Parallel Programming*, Mar. 2006.